Bidirectional analysis of the relationships between phenotypes and chemicals using public gene annotations

Christopher Hawthorne, MSc¹, Guillermo H. Lopez-Campos, PhD¹
¹Wellcome-Wolfson Institute for Experimental Medicine, Belfast, Northern Ireland, United Kingdom

Introduction

Almost fifteen years ago, Christopher P. Wild brought forward the idea of the exposome (all exposures that an individual encounters over the course of their life) with the intention of highlighting the necessity for methodologies to explore an individual’s exposome to a similar standard of genomic methods¹. Interest in the exposome concept has greatly developed in the last few years however compared to its genomic analog and the physical manifestations of the phenome it has yet to reach the same great heights that was envisioned by Wild. Biomedical informatics has greatly contributed to and enhanced the genomic and phenome research fields through ontologies such as the Gene Ontology (GO)²,³ which is regularly used for enrichment analysis to glimpse at the plausible biological roles of a set of genes, or through phenotype ontologies such as the Human Phenotype Ontology (HPO)⁴ respectively. Biomedical informatics is positioned to develop the tools necessary to advance exposomic research and deliver an interactivity framework between the exposome, genome and phenome⁵.

Currently, the exposome field has notable resources such as the Comparative Toxicogenomics Database (CTD)⁶ which provides curated links between chemicals and genes. CTD additionally offers links between chemicals and phenotypes which have been represented as GO terms. Although these approaches provide relevant insights, diseases are made up of different phenotypes whereas GO terms are focused on the biological outcomes at a cellular or molecular level. Therefore, we have detected the need to implement a method to relate chemical compounds to the different phenotypes that are encountered in human disease. For this purpose, we have developed phexpo (phenotype - exposome) a methodology to carry out bidirectional enrichment analysis of chemicals and phenotypes. This methodology has been bundled inside an R package. Phexpo incorporates chemical and gene data from CTD and phenotype and gene data from HPO. Using a chemical or phenotype derived gene list built from genes in both data sources, phexpo will provide enriched chemicals or phenotypes respectively.

Methods

Datasets of chemical and genes⁷, and chemical vocabulary⁸ were downloaded from the 5 February 2019 update of CTD. Phenotypes and gene annotations were downloaded from the 2019-02-12 version of HPO. The datasets used within phexpo were further preprocessed before integration. Phexpo is built and depends upon the following R packages; dplyr, tibble, ggplot2, magrittr, ranger, tidy and shiny.

Phexpo provides functions for the enrichment of single or multiple chemicals or phenotypes. The user inputs a selected chemical or phenotype into the respective function and phexpo by default will run a one-sided Fisher’s exact test to calculate enriched chemicals or phenotypes which are presented within a tibble data frame and arranged by p-value. To correct for multiple testing, phexpo employs Bonferroni correction and False Discovery Rate (FDR) p-value adjustments. Users can additionally visualise their results using a function which deploys a shiny interface.

Results

To demonstrate the capabilities of phexpo we present here two chemical analyses. One for the drug warfarin, used as a positive control, and chloroform as an industrial chemical example. We set a significant FDR threshold p-value <0.05. The enriched results are displayed using the visualisation function (Figure 1). To validate phexpo enrichment results we searched for and found evidence in the literature supporting these relationships. Warfarin’s enriched results mirror its application as a blood thinner, commonly prescribed to patients who suffer from deep vein thrombosis. Alternatively, the enriched results of chloroform show an association with lethargy and hepatic related phenotype terms (e.g. ‘Abnormality of the abdominal organs’, ‘Jaundice’, ‘Abnormality of the liver’ and ‘Abnormal liver morphology’). These phenotype terms reflect the known association of hepatotoxicity caused by chloroform⁹.
A - Warfarin

B - Chloroform

Figure 1. Enrichment result visualisation using visEnrich function incorporating a shiny app to show (A) the first nine warfarin results in tabular format or (B) chloroform bar chart results with enriched human phenotype ontology terms. In (B) p-value selection can be changed to original p-value, Bonferroni correction or FDR. Threshold value of p-value can additionally be chosen.

Conclusion

We are confident that phexpo provides a new approach for the analyses of chemical and phenotype interactions using enrichment analysis. This is a novel application in its bidirectional approach providing relationships from chemicals to phenotypes and from phenotypes to chemicals. Phexpo is not without its own limitations, although we utilise gold standard resources within the domain we only use one chemical-gene database and we are restricted to the information within the resources. Phexpo cannot detect the directionality of the association between a chemical and phenotype for example whether it is a treatment or a cause. Multiple chemicals or phenotypes analyses employ a simplified approach based in gene additivity and does not account for the magnitude or directionality of gene interactions. Phexpo is only for research purposes. Phexpo will be a valuable asset to further exposome research and we wish to keep expanding upon this initial version of phexpo.

References

7. Curated [chemical–gene interactions] data were retrieved from the Comparative Toxicogenomics Database (CTD), MDI Biological Laboratory, Salisbury Cove, Maine, and NC State University, Raleigh, North Carolina. World Wide Web (URL: http://ctdbase.org/). [February, 2019].
8. Curated [chemical vocabulary] data were retrieved from the Comparative Toxicogenomics Database (CTD), MDI Biological Laboratory, Salisbury Cove, Maine, and NC State University, Raleigh, North Carolina. World Wide Web (URL: http://ctdbase.org/). [February, 2019].